

Folding and unfolding of a triple-branch DNA molecule with four conformational states

Sandra Engel^{a,*}, Anna Alemany^{b,c}, Nuria Forns^{b,c},
Philipp Maass^{a,†}, and Felix Ritort^{b,c,‡}

^a*Fachbereich Physik, Universität Osnabrück,
Barbarastr. 7, 49076 Osnabrück, Germany*

^b*Departament de Física Fonamental, Facultat de Física,
Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain*

^c*CIBER-BBN Networking center on Bioengineering,
Biomaterials and Nanomedicine, Spain*

Abstract

Single-molecule experiments provide new insights into biological processes hitherto not accessible by measurements performed on bulk systems. We report on a study of the kinetics of a triple-branch DNA molecule with four conformational states by pulling experiments with optical tweezers and theoretical modelling. Three distinct force rips associated with different transitions between the conformational states are observed in the folding and unfolding trajectories. By applying transition rate theory to a free energy model of the molecule, probability distributions for the first rupture forces of the different transitions are calculated. Good agreement of the theoretical predictions with the experimental findings is achieved. Furthermore, due to our specific design of the molecule, we found a useful method to identify permanently frayed molecules by estimating the number of opened basepairs from the measured force jump values.

Keywords: Nonequilibrium systems, single-molecule experiments,
optical tweezers, DNA

PACS: 82.37.-j, 05.70.Ln, 82.39.Pj, 87.80.Nj

1 Introduction

In recent years, single-molecule experiments became of great importance in biophysical research since progress in nano- and microscale manufacturing technologies facilitated the design of scientific instruments with sufficient sensitivity and precision to enable the controlled manipulation of individual molecules (for reviews, see, for example, [22, 10]). In contrast to the traditionally used bulk assays, where individual biomolecular dynamics can get masked, single-molecule experiments provide new insights into

*sandra.engel@uos.de

†philipp.maass@uos.de, phone: 0049-541-969-3460, fax: 0049-541-969-2351,
<http://www.statphys.uni-osnabrueck.de>

‡fritort@gmail.com, phone: 0034-934035869, fax: 0034-934021149,
<http://www.ffn.ub.es/ritort/index.html>

the thermodynamics and kinetics of biophysical and biochemical processes hitherto not accessible. They complement standard spectroscopy and microscopy methods used in molecular biology and biochemistry and hence have to be regarded as an important source of additional information helping in the interpretation of biomolecular processes. Furthermore, single-molecule experiments permit the measurement of small energies and the detection of large fluctuations.

The manipulation of single molecules offers a powerful new tool in molecular and cellular biophysics allowing for the exploration of processes occurring inside the cell at an unprecedented level. To instance just a few of the recently investigated biochemical processes: the transport of matter through pores or channels [19, 20, 1], interactions between DNA and proteins [13] or DNA and RNA [36], the motion of single-molecular motors [2, 4, 32], DNA transcription and replication [34, 15], virus infection [26, 8], DNA condensation [23] and ATP generation [35]. In addition, the structure of biological networks [30] and the viscoelastic and rheological properties of the DNA [27, 31, 6, 29] have been studied.

An important class of single-molecule experiments are performed with optical tweezers. By means of an optical trap generated by a focused laser beam, this useful technique renders it possible to exert forces on micron sized objects, achieving sub-piconewton and sub-nanometer resolution in force and extension, respectively. Accordingly, one can study force-induced folding-unfolding dynamics and in this way get insight into corresponding processes in the cell and typical bond forces. Of particular interest is the unfolding of DNA molecules, where the hydrogen bonds between the complementary base pairs (bps) are disrupted. This so-called unzipping is connected to the DNA replication mechanism.

An interesting field of biophysical studies is the investigation of junctions in molecules since they present manifold ways to interact with other substances, for instance cations. Three-way junctions are especially interesting because metal ions such as magnesium can bind to them and alter the tertiary structure. Here a first step of such a study is presented where we investigate a molecule with a three-way junction alone, without cation binding.

Many of the single-molecule experiments so far focused on molecules with a relatively simple free energy landscape (FEL) exhibiting just two states, a folded and an unfolded one, or including an additional misfolding state, leading to different kinetic pathways. In this work we will consider a richer situation, where metastable states as intermediates occur during the folding-unfolding route. In this context, we will address the following key questions:

- (i) Can a corresponding molecule with such intermediate states be designed on the basis of a suitable model for a FEL?
- (ii) Is it possible to observe the intermediate states by performing pulling experiments with optical tweezers?
- (iii) Can phenomenological Bell-Evans kinetic models be applied to describe the folding-unfolding processes including intermediate states? In particular, when validating the kinetic theory against the experimental results, how do the first rupture force distributions compare with the ones predicted by the theory?

A further important aspect that we looked at in some detail is the heterogeneity of molecular folding-unfolding behaviour that we observed in the experiments. By measuring force-distance curves (FDCs) of several molecules we classified them into

different reproducible patterns. This leads to a useful method to identify irreversible molecular fraying, a phenomenon which is often observed in single-molecule studies.

2 Description of the experiments

Based on Mfold folding predictions [25, 37] and taking FEL considerations into account (see sec. 5), we designed and synthesised a DNA molecule which is composed of three parts and hence referred to as triple-branch molecule. It consists of a stem as introduced in [21] with 21 bps and two nearly identical hairpin branches which are formed of 16 bps and a loop with four bases, thus comprising a total number of 114 bases, see fig. 1. To avoid misfolding, the second hairpin branch differs at two positions from the first one.

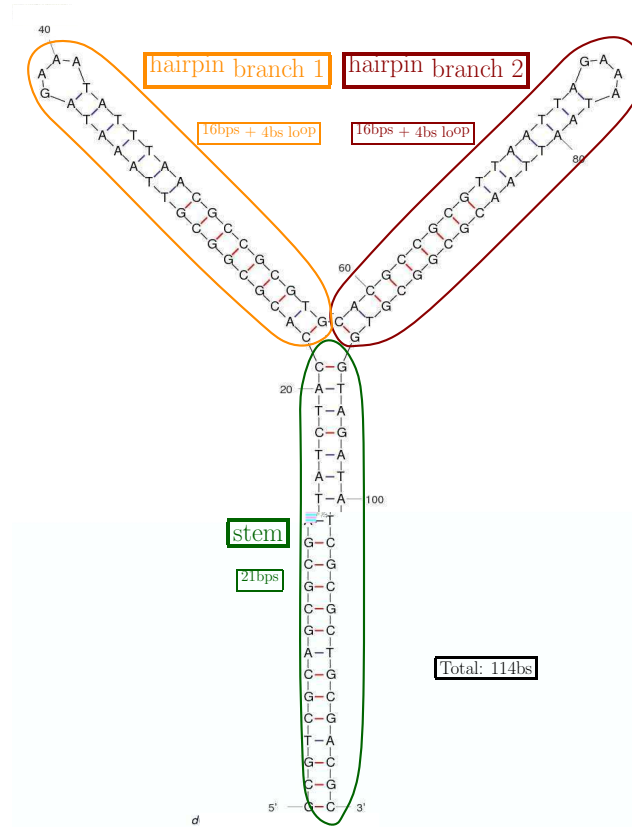


Figure 1: Structure of the triple-branch DNA molecule.

The triple-branch molecule is inserted between two identical short double-stranded DNA (dsDNA) handles of 29 bps each [9], leading to a total number of 172 bases, corresponding to a total contour length in the unfolded state of about 100 nm. Each of these polymer spacers is chemically linked to a bead. One of the beads is retained with the help of a pipette via air suction, the other one is optically trapped in a laser focus [28]. On the 5' end of the DNA molecule, biotin is attached to enable a connection with

a streptavidin-coated bead (SA bead) whose diameter is $1.8\text{ }\mu\text{m}$. Biotin is a vitamin which establishes a strong linkage to the proteins avidin and streptavidin. The 3' end is modified with the antigen digoxigenin, able to interact with an antidigoxigenin-coated bead (AD bead). The latter has a diameter of $3.0\text{ }\mu\text{m}$. Figure 2 shows the different components of the molecular construct, that is to say the triple-branch DNA molecule, the handles and the beads, captured in optical trap and micropipette, respectively. Note that it is not a true-to-scale representation.

The pulling experiments are carried out with a miniaturised dual-beam laser optical tweezers apparatus [11] at room temperature ($\approx 25\text{ }^\circ\text{C}$) and at salt concentration of 1 M NaCl aqueous buffer with neutral pH (7.5) stabilised by Tris HCl and 1 M EDTA. The dual-beam optical tweezers collect data at 4 kHz and can operate with a feedback rate of 1 kHz. Spatial resolution constitutes 0.5 nm with a maximal distance range of $\approx 10\text{ }\mu\text{m}$. Forces up to 100 pN can be achieved, whereas the force resolution is 0.05 pN.

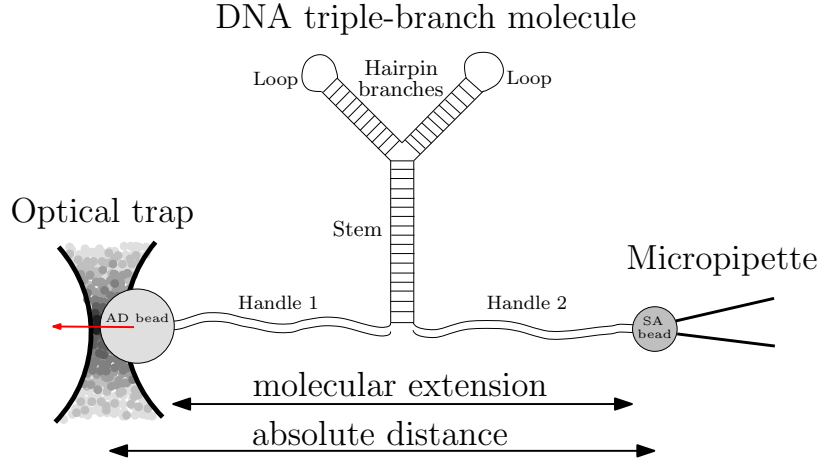


Figure 2: Sketch of the experimental setup. In our experiments we measure the relative distance X rather than its absolute value.

Optical tweezer pulling experiments permit the measurement of the force f as well as the total distance X between the centre of the optical trap and the tip of the micropipette, see fig. 2. In the experiment we vary the trap-pipette distance $X(t)$ with a constant speed $v = dX/dt$ in the range of 45 nm/s to 200 nm/s, which corresponds to a constant average loading rate r of 3.0 pN/s and 13.4 pN/s in between rip events, respectively. The experiment consists of loading cycles which in turn are divided into an unfolding part (during "pulling") and a folding part (during "pushing"). The loading cycles are repeated as long as the tether connection is unbroken. Otherwise a new connection has to be established, possibly a new molecule must be searched and linked to a new bead. In sec. 4 we will discuss different patterns found in the measured curves and present a detailed analysis of two representative molecules. We chose them among seven molecules exhibiting the first and among five molecules featuring the second pattern. For each molecule we recorded, on average, approximately 50 cycles. The pulling speeds, ranging from 45 to 200 nm/s, influence the experimental results only weakly due to a logarithmic dependence of the first rupture force with the speed. We found compatible data for sets of similar molecules. In the theoretical analysis of the data in sec. 5, we concentrate on the largest set of 82 loading cycles for the molecule

at a speed of 200 nm/s. The study of the above mentioned second molecule comprises 55 cycles.

3 Analysis of unfolding and folding trajectories

Based on the design of the triple-branch molecule, we have to distinguish between four conformational states (see fig. 3):

1. a completely folded molecule.
2. a completely unfolded stem with the hairpin branches still folded.
3. stem and either hairpin branch 1 or 2 are completely unfolded.
4. a completely unfolded molecule.

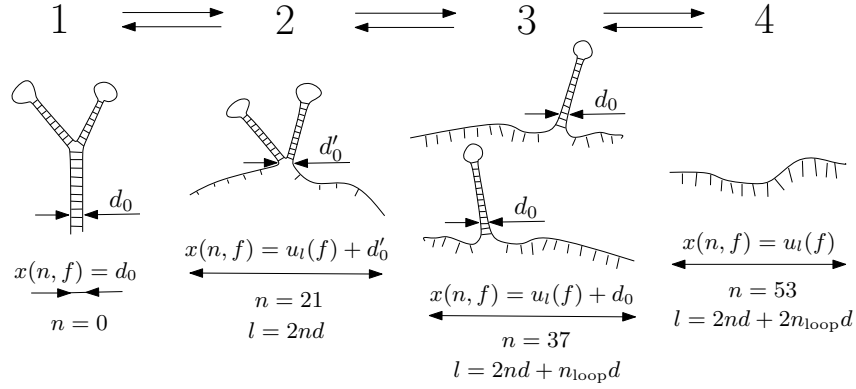


Figure 3: The four stable or metastable states of the triple-branch molecule: 1 - folded molecule, 2 - unfolded stem, 3 - stem and one hairpin branch unfolded, 4 - unfolded molecule. The values of $x(n, f)$ refer to the end-to-end distance given in eq. (2) with the number n of opened bps corresponding to the molecular construct shown in fig. 1 and the contour length l according to eq. (3).

Figure 4(a) displays a typical unfolding route and fig. 4(b) a typical unfolding and folding trajectory in form of a FDC, which in fact records the evolution of the force as a function of time and relative trap position X ¹. With rising X , the force f first increases almost linearly according to an elastic response of the DNA handles, which consist of dsDNA and are stable over the whole range of forces where the unfolding / folding of the triple-branch molecule takes place. The overstretching transition of the linkers, typically at 65 pN, lies much above the forces we explore. At a first rupture force f_1 a sudden decrease ("jump") Δf_1 occurs, which is caused by the unfolding of the stem. This unfolding goes along with an abrupt change in the length of the molecule when single-stranded DNA (ssDNA) is released. As a consequence, the bead in the optical

¹ In our experiments we measure the relative distance between trap and pipette, X , rather than the absolute value. The force f exerted on the molecular construct leads to a displacement f/k_b of the bead in the optical trap, where $k_b = 0.08$ pN/nm is the rigidity of the trap. Hence, the relative distance X is related to the relative molecular extension x_m (see fig. 2) by $x_m = X - f/k_b$.

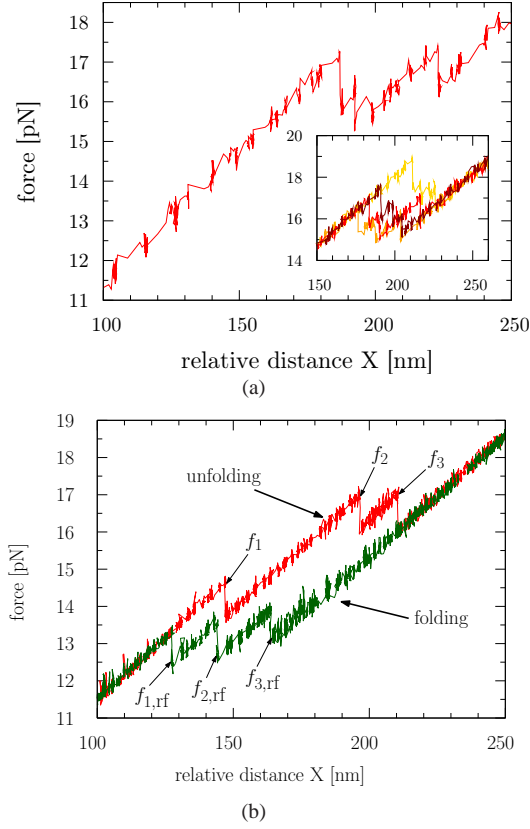


Figure 4: Force as a function of change in the trap-pipette distance (a) for one typical unfolding trajectory for the first investigated molecule and (b) for one typical unfolding and folding trajectory for the second investigated triple-branch molecule. Indicated are the three first rupture forces belonging to the transitions 1 (f_1), 2 (f_2) and 3 (f_3) and the respective refolding forces, labelled $f_{i,rf}$. The inset in (a) shows four further unfolding curves for the first investigated molecule.

trap moves towards the centre of the trap, visible as the force drops. Following the jump Δf_1 , there is again a linear increase up to the next force rip at a first rupture force f_2 , where one of the hairpin branches unfolds, which in turn leads to the force jump Δf_2 . Eventually, the second hairpin branch unzips at a first rupture force f_3 with a jump Δf_3 . The linear regime following this last force rip corresponds to the stretching of the whole molecular construct including handles and the already unfolded triple-branch molecule.

Upon decreasing X from the completely unfolded state 4, the force first follows closely the corresponding unfolding part of the trajectory. However, a backward transition does not occur at $(f_3 - \Delta f_3) \approx 16.2$ pN, but at a considerably lower value of (14.0 ± 0.8) pN, cp. fig. 4(b), hence manifesting a hysteresis effect. Moreover, an investigation of a larger number of folding trajectories reveals that the bases of the hairpin branches do not always pair conjointly in well-defined events during a short time inter-

val. In contrast, the corresponding force rips during unfolding indicate a cooperative behaviour of the biomolecule, where the breakage of all hydrogen bonds stabilising the DNA structure happens almost simultaneously. The second transition takes place at (13.3 ± 0.8) pN instead of $(f_2 - \Delta f_2) \simeq 15.5$ pN. After both hairpin branches refolded, one can identify another sharp transition to the folded state 1 around (11.2 ± 0.8) pN, which is again considerably lower than $(f_1 - \Delta f_1) \simeq 13.7$ pN.

During unfolding, one can assume that the breakage of hydrogen bonds follows the sequence of the molecular construct. This makes it useful to introduce the number n of broken bonds as state variable and to calculate a FEL as function of this variable (see sec. 5). The refolding of the hairpin branches in the folding trajectories exhibit less sharp transitions, see fig. 4(b). During folding, in particular at the beginning in the unfolded state, a huge number of secondary structures can be found which implies that the kinetic pathways are less predefined and accordingly, the transitions get smeared out. With respect to a theoretical treatment, moreover, a description in terms of the simple state variable n becomes unlikely to be sufficient. In a refined analysis, many more configurations should have to be included as relevant states in a coarse-grained description [17]. Such refined analysis, however, goes beyond the scope of this work and we therefore concentrate on the unfolding process in the following.

There are plenty of possible ways to analyse the FDCs in order to find out the first rupture forces and the force jumps. In our procedure we arranged the normalised data, i.e. the relative distance X and the force f , in windows of a certain size of data points. For all consecutive windows we then calculated the slope of the considered data points, the span, i.e. the maximum distance between the lowest and the highest force value, and the mean of the relative distance \bar{X}_j as a moving average. Transitions between the conformational states take place where the slope is minimal and the span maximal under the condition that an appropriate number of contiguous windows is connected. Having found the \bar{X}_j of the three force rips, slope and axis intercept are calculated by linear regression for each conformational state. One can now easily calculate the first rupture forces as the intersection points with the four fitted lines and extract the force jump values, as exemplified in fig. 5 for both molecules whose unfolding trajectories were depicted in figs. 4(a) and 4(b), respectively. Since the data acquisition rate is

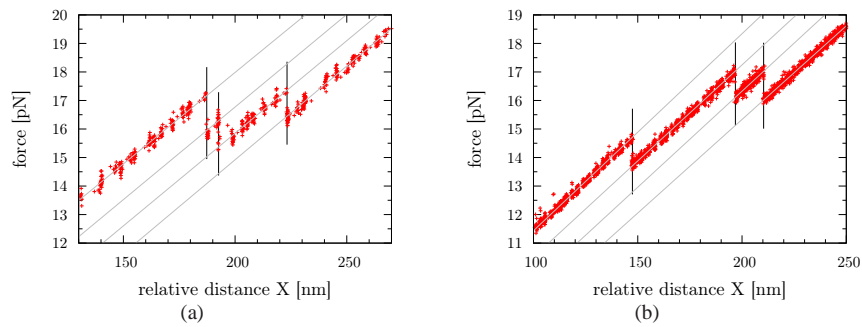


Figure 5: The first rupture forces and force jump values of the unfolding trajectories shown in fig. 4 are extracted as indicated here from the intersection of the fitted grey lines and the transitions (vertical lines). In part (a) the procedure is shown for the molecule used in the theoretical analysis in sec. 5 and in (b) for the molecule with permanent fraying behaviour.

constant, at higher pulling speeds less data points are collected and therefore the values of the analysis are broader distributed. The three first rupture forces f_i as well as the jumps Δf_i are subject to stochastic fluctuations, as can be seen in fig. 4(a) where we show four unfolding trajectories belonging to different pulling cycles of the same molecule in the inset. An analysis revealed that the fluctuations of the force jumps Δf_i are about ten times smaller than the fluctuations of the f_i . Accordingly, in sec. 5, we will disregard the fluctuations in the Δf_i and use only their averages $\overline{\Delta f_i}$ that will be discussed in more detail in the following section.

4 Different unfolding patterns related to the number of opened bps

Applying the above mentioned procedure to analyse the experimental data of several molecules, we found two predominating patterns in the unfolding trajectories which are reflected in the distributions of the first rupture forces f_i as follows. In the first pattern, see fig. 6(a), these distributions have a similar shape for all three force rips. The histograms indicate the existence of one maximum slightly below 17 pN. The mean values for the three first rupture forces are $\overline{f_1} = (17.0 \pm 0.9)$ pN, $\overline{f_2} = (16.4 \pm 0.6)$ pN and $\overline{f_3} = (16.8 \pm 0.7)$ pN. In contrast, in a second pattern we detected a strikingly lower value for the first rupture force of the first rip $\overline{f_1} = (14.6 \pm 0.8)$ pN, as depicted in fig. 6(b), whereas the other two rip forces lie basically in the same range of 16 to 18 pN. As in the former case, the second rip tends to have a slightly smaller first rupture force, $\overline{f_2} = (16.5 \pm 0.8)$ pN, than the third rip, $\overline{f_3} = (17.2 \pm 0.5)$ pN.

The small $\overline{f_1}$ observed in the second molecule suggests the occurrence of permanent molecular fraying. Obviously less force is needed to unfold the stem than typically, cp. fig. 4(a) with 4(b), since some bps at its basis are partly or completely melted. In other words, during folding, this molecule does not reach an entirely folded state but some bps of the stem next to the handles remain irreversibly and permanently open. This phenomenon has been observed previously in several pulling experiments [11, 33]. A possible reason for irreversible fraying is the formation of reactive oxidative species due to the impact of the laser light of the optical trap leading to a degradation of the DNA bases [12]. The so generated singlet oxygens are known to oxidise certain nucleic acids, such as guanine and thymine, irreversibly. This could explain our observation that once a molecule shows fraying it does not change back again to normal behaviour. Due to the fact that we work with polystyrene microspheres which are more prone to photodamage than the DNA bases themselves, their wide ranging interaction with the bases might be reduced replacing polystyrene by silica beads, which exhibit considerably minor irreversible oxidative damage. It would be very interesting to carry out such experiments.

The insets in figs. 6(a) and 6(b) depict the corresponding force jump distributions of both molecules. While the first molecule possesses a large first force jump value of $\overline{\Delta f_1} = (1.3 \pm 0.2)$ pN and two smaller force jumps at the second and third rip of $\overline{\Delta f_2} = (0.9 \pm 0.2)$ pN and $\overline{\Delta f_3} = (1.0 \pm 0.2)$ pN, the frayed molecule features three force jumps of approximately the same value, i.e. $\overline{\Delta f_1} = (0.93 \pm 0.05)$ pN, $\overline{\Delta f_2} = (0.95 \pm 0.06)$ pN and $\overline{\Delta f_3} = (0.97 \pm 0.04)$ pN, respectively. This illustrates clearly the influence of irreversible fraying in the latter case since, roughly estimated, the same number of bps is expected to open in all three rips which should not be the case in an entirely folded molecule.

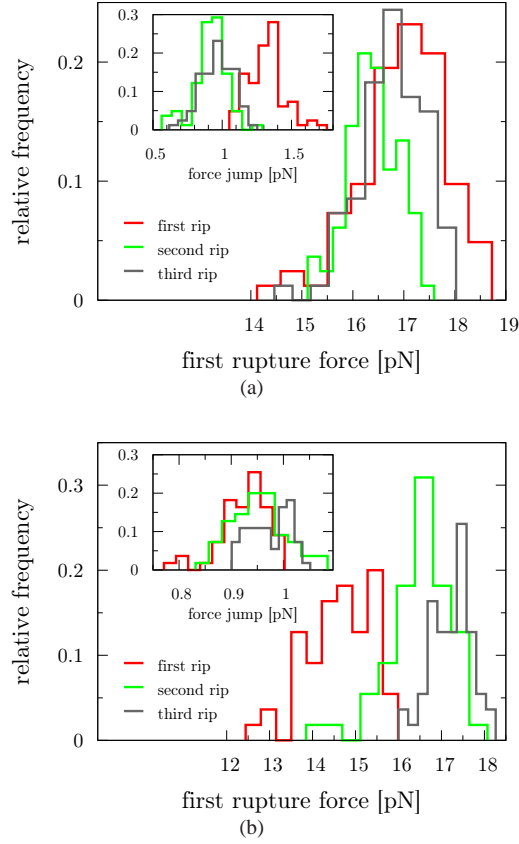


Figure 6: Histograms of the three first rupture forces during unfolding for two representative triple-branch molecules: (a) the one used in the theoretical analysis in sec. 5 and (b) the one exhibiting permanent fraying behaviour. Note that the average force value of the first rip has decreased in (b) as compared to (a). The insets show the corresponding histograms for the force jump values.

During a force rip, the relative distance X is constant and thus $\Delta X = 0$. Therefore the change in the relative molecular extension,

$$\Delta x_m = \Delta X - \Delta f / k_{\text{eff}} = \Delta x(n, f), \quad (1)$$

is only related to the force jump Δf and the combined stiffness of bead and handles k_{eff} , given by $1/k_{\text{eff}} = 1/k_b + 1/k_h$, where k_b is the trap stiffness and k_h the rigidity of the handles, respectively. With the help of a linear least squares fit, the effective stiffness k_{eff} of the molecular construct is extracted from the average slope of the FDCs and amounts to (0.067 ± 0.003) pN/nm. Note that, as expected, k_{eff} is smaller than k_b (≈ 0.08 pN/nm).

For a certain force value f and assuming an elastic model for the released ssDNA, the number n of opened bps is related univocally to the equilibrium end-to-end distance of the DNA molecule $x(n, f)$, whereas its change $\Delta x(n, f)$, in turn, equals Δx_m . Using this relation, it is now possible to estimate the change in the number Δn_i of bps which

are opened sequentially during each force rip. Considering only the configurations of the four conformational states, the equilibrium end-to-end distance $x(n, f)$ can be decomposed into two parts, cp. fig. 3. The first part, the elongation $u_l(f)$ of the mean end-to-end distance of the ssDNA along the force direction, accounts for the ideal elastic response of the ssDNA, where l is the contour length. The second part contains the contribution of the diameter of stem and hairpin branches, respectively. Accordingly,

$$x(n, f) = u_l(f) + \begin{cases} 0, & n = 21 + 16 + 16 \text{ bps} \\ d_0, & n = 0 \text{ and } 21 + 16 \text{ bps} \\ d'_0, & n = 21 \text{ bps} \end{cases}, \quad (2)$$

where $d_0 \simeq 2 \text{ nm}$, in accordance with the diameter of the B-DNA helix. The exact value of the diameter contribution of both hairpin branches d'_0 , when the stem is unfolded, depends on the orientation of the branches. We set $d'_0 = 2d_0$ as a working value².

Regarding the contour length l , which depends, amongst others (cp. sec. 5), on the number n of opened bps, and considering again solely the configurations of the four conformational states, one gets

$$l = 2nd + \begin{cases} 0, & n = 0 \text{ and } 21 \text{ bps} \\ n_{\text{loop}}d, & n = 21 + 16 \text{ bps} \\ 2n_{\text{loop}}d, & n = 21 + 16 + 16 \text{ bps} \end{cases}, \quad (3)$$

where the interphosphate distance d is taken to be 0.59 nm/base and $n_{\text{loop}} = 4$ is the number of bases per loop. For every force rip f_i the corresponding number of opened bps n_i is calculated separately by considering the differences of l between the states.

Different types of models can be used to calculate $u_l(f)$. Prominent examples borrowed from polymer physics are the freely jointed chain (FJC) and the worm-like chain (WLC) model. According to ref. [27], the FJC model includes an extra term, leading to the expression

$$u_l(f) = l \left(1 + \frac{f}{Y} \right) \left[\coth \left(\frac{bf}{k_B T} \right) - \frac{k_B T}{bf} \right]. \quad (4)$$

Here Y denotes the Young modulus, b is the Kuhn length, k_B the Boltzmann constant and T the temperature. Typical values of the Kuhn length and the Young modulus under working conditions of $T \simeq 25 \text{ }^\circ\text{C}$ and 1 M NaCl concentration are $b = 1.42 \text{ nm}$ and $Y = 812 \text{ pN}$ [27] or, as published recently, $b = 1.15 \text{ nm}$ and $Y = \infty$ [11], respectively. In the WLC model [3], the force $f(u_l)$, due to an elongation u_l , is given by

$$f(u_l) = \frac{k_B T}{P} \left[\frac{1}{4(1 - u_l/l)^2} - \frac{1}{4} + \frac{u_l}{l} \right], \quad (5)$$

and to obtain $u_l(f)$, this equation has to be inverted. Based on the WLC model, we tested the influence of the persistence length P in a typical range of 1.0 to 1.5 nm [6].

From the data shown in table 1 it is apparent that, depending on the model and parameters, the results for the estimated change in the number of opened bps vary in an acceptable range when the errors are taken into account. In addition, one can see that it is not evident which model and parameters should be considered as the best ones. Good results are found for the FJC model using recent values of [11] and for the WLC model

²Due to this simplification, the Δn_1 of the first rip is likely to be slightly underestimated and the second rip's Δn_2 overestimated. However, it will not affect the change in the total number Δn_{tot} of opened bps since we consider the change of $x(n, f)$, and the d'_0 contributions will cancel each other out.

		molecule 1				molecule 2			
model	parameter b & P [nm], Y [pN]	change in no. of opened bps [bps]							
		Δn_1	Δn_2	Δn_3	Δn_{tot}	Δn_1	Δn_2	Δn_3	Δn_{tot}
FJC	$b = 1.42, Y = 812$ [27]	18 (2)	14 (2)	15 (2)	46 (3)	13 (1)	15 (1)	15 (1)	42 (2)
	$b = 1.15, Y = \infty$ [11]	19 (3)	15 (2)	16 (2)	50 (3)	14 (1)	16 (1)	16 (1)	46 (2)
WLC	$P = 1.0$	20 (3)	16 (3)	17 (3)	54 (3)	14 (1)	17 (1)	17 (1)	49 (2)
	$P = 1.3$ [24]	20 (3)	15 (2)	16 (2)	51 (3)	14 (1)	16 (1)	16 (1)	46 (2)
	$P = 1.5$	19 (3)	15 (2)	16 (2)	50 (3)	13 (1)	16 (1)	16 (1)	45 (2)
expected values		21	16	16	53	21	16	16	53

Table 1: Overview over the change in the number of opened bps for different models and parameters for the two representative molecules. The numbers in brackets are the standard deviations.

with $P = 1.3$ nm [24]. We chose to work with the FJC model with the parameters of [11] in sec. 5.

The second molecule indeed reveals a considerably smaller Δn_1 , depending on the model around 13 or 14 bps, so that 7 or 8 bps are not closed after the folding process is completed. Performing single-molecule experiments without knowing the exact influence of permanently frayed bps can lead to misinterpreted results. Checking the appropriate parameters for the polymer models with the help of the change in the number of opened bps of the hairpin branches, one can estimate the number of irreversibly frayed bps at the basis of the stem³.

5 Theory for the unfolding kinetics

The kinetics of the unfolding process can be described on a coarse-grained level based on a Gibbs free energy $G(n, f)$ as a function of the number n of sequentially opened bps for an applied force f . For small forces, including $f = 0$, the FEL is expected to have a shape as displayed in fig. 7(a). In general, $G(n, 0)$ increases monotonously with n . However, local minima occur at the metastable states 2, 3 and 4 because there is an increase of entropy associated with the release of additional degrees of freedom when the stem-hairpin-junction and end-loops of the hairpin branches are opened. With rising force the FEL is expected to get tilted, so that the energies of the metastable states are lowered. With a knowledge of $G(n, f)$ we can apply standard transition rate theory and write for the transition rate from state i to $i + 1$

$$\Gamma_{i,i+1}(f) = \gamma_i^0 \gamma_{i,i+1}(f), \quad (6)$$

where γ_i^0 is an attempt rate and $\gamma_{i,i+1}(f)$ is the Boltzmann factor corresponding to the activation barrier $\Delta G_{i,i+1}(f)$ that has to be surmounted,

$$\gamma_{i,i+1}(f) = \exp\left(-\frac{\Delta G_{i,i+1}(f)}{k_B T}\right). \quad (7)$$

When considering only sequential configurations in the evaluation of the FEL, different structures compatible with a given n can occur once the stem is completely unfolded. These refer to different possibilities of breaking the bps in the hairpin branches

³We like to note that the checking of the change in the number of opened bps can be, in principle, also applied to non-permanent, reversible molecular fraying.

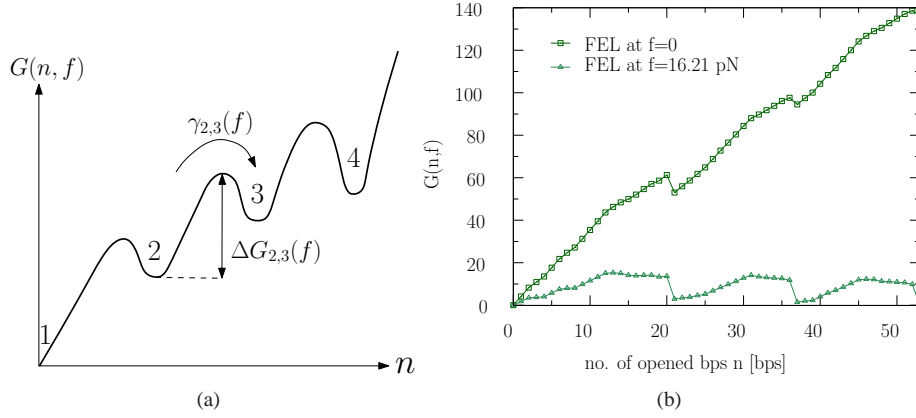


Figure 7: (a) Sketch of the expected FEL for the triple-branch molecule at zero force as a function of the number n of opened bps. (b) FEL calculated from eqs. (8) to (11) at two different forces in units of the thermal energy $k_B T$.

1 and 2. In order to point out this "degeneration", a new parameter α is introduced, leading to the FEL $G(n, \alpha, f)$. At a given n , we must calculate the (restricted) partition sum over the configurations α to get $G(n, f)$. In order to find $G(n, \alpha, f)$, we consider the following decomposition,

$$G(n, \alpha, f) = G_{\text{form}}(n, \alpha) + G_{\text{str}}^{\text{ss}}(n, \alpha, f) - f \Delta x_l^{\text{ss}}, \quad (8)$$

where $G_{\text{form}}(n, \alpha)$ is the free energy of formation of the configuration (n, α) , $G_{\text{str}}^{\text{ss}}(n, \alpha, f)$ is the strain energy of the unfolded ssDNA and $f \Delta x_l^{\text{ss}}$ is a Legendre term (for the definition of Δx_l^{ss} see eq. (10) below).

The free energy of formation $G_{\text{form}}(n, \alpha)$ is written as

$$G_{\text{form}}(n, \alpha) = \sum_{\text{all bps}} g_{\mu, \mu+1} + G_{\text{junc}}(n) + G_{\text{loop}}^{(1)}(n) + G_{\text{loop}}^{(2)}(n). \quad (9)$$

The first term refers to the nearest neighbour model developed in [5, 7], which specifies the interaction $g_{\mu, \mu+1}$ between a base pair μ and the directly adjacent one $\mu + 1$. It was shown to provide reasonable agreement with experiments [16, 21, 33]. For example, applying this model onto a sequence 5'-TCCAG...-3' and its complementary part 3'-AGGTC...-5', the stack energy reads $G_{\text{stack}} = g_{\text{TC/AG}} + g_{\text{CC/GG}} + g_{\text{CA/GT}} + g_{\text{AG/TC}} + \dots$. The most recent values of $g_{\mu, \mu+1}$ lie in the range of -2.37 to -0.84 kcal/mol at 25°C [11]. The terms $G_{\text{junc}}(n)$, $G_{\text{loop}}^{(1)}(n)$ and $G_{\text{loop}}^{(2)}(n)$ in eq. (9) describe the free energy reduction due to the release of the stem-hairpin-junction and end-loops and are estimated from [25, 37] as $G_{\text{loop}}^{(1)} = G_{\text{loop}}^{(2)} = 1.58$ kcal/mol and $G_{\text{junc}} = 4.90$ kcal/mol.

The strain energy $G_{\text{str}}^{\text{ss}}(n, \alpha, f)$ of the unfolded single-stranded part [21] with contour length $l = l(n, \alpha)^4$ can be calculated from the work needed to stretch the unpaired bases. We like to remind the reader that we denote the elongation of the mean end-to-end distance of the ssDNA in force direction by $u_l(f)$. In what follows we chose

⁴In previous publications of some of the authors this contour length was denoted by $l_{n, \alpha}$ to emphasize the dependence on n (and, in addition, α here). This dependence is caused by the change of the contour length in the transitions. For easier reading we suppress to give it explicitly in the following. Further details about the contour length were already discussed in sec. 4.

to work with the FJC model, see eq. (4), with parameters of [11] instead of using the WLC model, despite the fact that both approaches give similar good results (cp. table 1). Since $u_l(f)$ is monotonously increasing with f , it has an inverse $f_l(u) = u_l^{-1}(f)$, which is the force that is exerted by a ssDNA chain with contour length l , if its mean end-to-end distance is elongated by u . Accordingly, setting $\Delta x_l^{\text{ss}} = u_l(f)$, we can write

$$G_{\text{sr}}^{\text{ss}}(n, \alpha, f) = \int_0^{\Delta x_l^{\text{ss}}} du' f_l(u') = f \Delta x_l^{\text{ss}} - \int_0^f df' u_l(f'). \quad (10)$$

Finally, we computed $G(n, f)$ by

$$G(n, f) = -k_B T \ln \sum_{\alpha} \exp\left(-\frac{G(n, \alpha, f)}{k_B T}\right). \quad (11)$$

In fig. 7(b) the FEL is depicted for $f = 0$ and $f = 16.21$ pN. It exhibits the behaviour anticipated in fig. 7(a): for zero force it has minima at the stable/metastable states and it becomes tilted with rising force. When approaching the force regime where the rips occur in fig. 4(a), the levels of the minima become comparable. We want to point out that the DNA sequences shown in fig. 1 have been designed deliberately to yield the multiple-state structure seen in fig. 7(b). This gives us some confidence in the model underlying the construction of the $G(n, \alpha, f)$ in eq. (8).

Based on the FEL we can easily calculate the transition probability $W(f_i|f_{i-1})$ for the first rupture force f_i in the i th transition if the first rupture force was f_{i-1} in the $(i-1)$ th transition. The result is

$$W(f_i|f_{i-1}) = \frac{\gamma_i^0}{r} \gamma_{i,i+1}(f) \exp\left[-\frac{\gamma_i^0}{r} \int_{f_{i-1}}^f df' \gamma_{i,i+1}(f')\right], \quad (12)$$

where $f_i^* = f_i - \overline{\Delta f_i}$ (and $f_0 = f_0^* = 0$); r was the loading rate, see sec. 2, and γ_i^0 and $\gamma_{i,i+1}(f)$ were defined in eq. (6). The activation energy $\Delta G_{i,i+1}(f)$ appearing in eq. (7) was calculated, as indicated in fig. 7(a), from the $G(n, f)$ by determining the energy $G_i^{\text{min}}(f)$ of the local minimum belonging to state i and the saddle point energy $G_{i,i+1}^{\text{saddle}}(f)$ of the i th transition between the i th and $(i+1)$ th state, $\Delta G_{i,i+1}(f) = G_{i,i+1}^{\text{saddle}}(f) - G_i^{\text{min}}(f)$. The attempt rate γ_i^0 was used as the only fitting parameter.

For the joint probability density of the three first rupture forces we then obtain

$$\Psi_3(f_1, f_2, f_3) = W(f_1|0) W(f_2|f_1) W(f_3|f_2), \quad (13)$$

which allows us to calculate the distributions shown in fig. 6(a).

Figure 8 displays the histograms for the three rips from fig. 6(a) in comparison with the distributions calculated from our theory. In view of the available statistics (82 cycles, see sec. 2), the agreement is quite satisfactory.

6 Conclusions

An important class of biophysical studies is the investigation of junctions in molecules since they present manifold ways to interact with other substances, for instance cations. Three-way junctions are especially interesting because metal ions such as magnesium [14, 18] can bind to them and alter the tertiary structure. Here a first step of such a

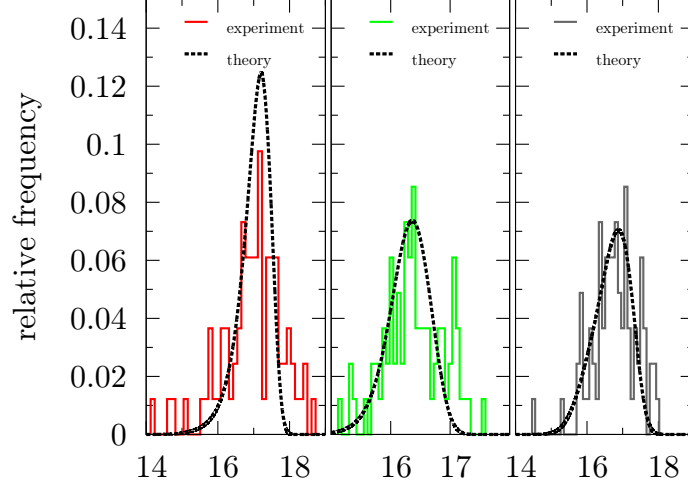


Figure 8: Comparison of the first rupture force distributions from fig. 6(a) with the theory for the first (f_1), the second (f_2) and the third (f_3) transition, see left, middle and right panel, respectively. The best fit was obtained for attempt rates $\gamma_1^0 = 9 \cdot 10^6$ Hz, $\gamma_2^0 = 1.1 \cdot 10^6$ Hz and $\gamma_3^0 = 3.5 \cdot 10^5$ Hz.

study is presented where we investigate a molecule with a three-way junction alone, without in vivo relevant substances.

One of our aims of this work was to study whether the construction and kinetics of more complex DNA molecules with richer folding-unfolding behaviour can be described by proper extensions of theories developed successfully for two-state systems so far. Our results show that this is indeed possible, at least for the unfolding trajectories. A triple-branch molecule has been specifically designed to produce a four-state system based on a model for the free energy landscape. This design was successful and we were able to prove the existence of these states by the emergence of associated force rips in pulling experiments. The first rupture forces have been systematically recorded in these pulling experiments and their distributions have been calculated. A transition rate theory based on the free energy landscape was successful in describing these distributions.

Two patterns have been found in the measured unfolding trajectories, one indicating the anticipated unfolding behaviour and the other one pointing to the occurrence of irreversible molecular fraying. This characterisation was possible by connecting the extracted force jump values to the change in the number of opened bps at each transition. For this estimation we tested the validity of two polymer models for the elastic response of ssDNA (FJC and WLC) and different sets of parameters in order to find the best agreement with the expected values. This analysis is useful to compare the elastic properties measured in DNA unzipping experiments with those obtained by stretching ssDNA polymers [11].

One class of molecules required a smaller force than anticipated to unfold the stem since some bps at its basis are partly or completely melted due to photodamaging. Permanent molecular fraying is an usually undesired, but frequent effect in single-

molecule studies and deserves special attention in order to reduce its distorting influence on experimental results. It is important to find means to avoid irreversible fraying since not fully closed molecules change the measured unfolding-folding trajectories so that an average over all molecules, including permanently frayed ones, can lead to deviations of the real values and to misinterpreted results. To improve the statistics of the results, it is therefore necessary to identify irreversibly frayed molecules and remove them from the analysis. Within our analysis, we found a useful method to identify permanent molecular fraying. With the appropriate parameters for the polymer models one can estimate the number of irreversibly frayed bps at the basis of the stem. Further experiments performed with silica beads instead of polystyrene microspheres could clarify under which conditions permanent molecular fraying can be decreased.

All this knowledge paves the way to further interesting studies such as the refolding problem [17] or two topics which require further experimental research. Firstly, the binding of metal ions to the three-way junctions could be examined in order to find out about structural changes due to the formation of tertiary contacts and therefore altered kinetics of the unfolding process [18]. Secondly, the translocation motion of helicases that unwind dsDNA could be addressed as another interesting subject, including the investigation of how they move along bifurcation points. Eventually, the kinetic approach for the prediction of the probability distributions of the first rupture forces could be a well suited starting point for future in-depth modelling in this domain on a more microscopic basis.

Acknowledgements

S. E. thanks the Deutscher Akademischer Austauschdienst (DAAD) for providing financial support (FREE MOVER and PROMOS) for stays at the Small Biosystems Lab in Barcelona where the experiments have been performed. A. A. is supported by grant AP2007-00995. F. R. is supported by the grants FIS2007-3454, Icrea Academia 2008 and HFSP (RGP55-2008).

References

- [1] Ammenti, A., F. Cecconi, U. M. B. Marconi, and A. Vulpiani. 2009. *J Phys Chem B* **113** (30):10348–10356.
- [2] Block, S. M., C. L. Asbury, J. W. Shaevitz, and M. J. Lang. 2003. *Proc Natl Acad Sci U S A* **100** (5):2351–2356.
- [3] Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. *Science* **265** (5178):1599–1600.
- [4] Carter, N. J., and R. A. Cross. 2006. *Curr Opin Cell Biol* **18** (1):61–67.
- [5] Crothers, D. M., and B. H. Zimm. 1964. *J Mol Biol* **9**:1–9.
- [6] Dessinges, M.-N., B. Maier, Y. Zhang, M. Peliti, D. Bensimon, and V. Croquette. 2002. *Phys Rev Lett* **89** (24):248102.
- [7] Devoe, H., and I. J. Tinoco. 1962. *J Mol Biol* **4**:500–517.

- [8] **Dumont, S., W. Cheng, V. Serebrov, R. K. Beran, I. Tinoco, A. M. Pyle, and C. Bustamante.** 2006. *Nature* **439** (7072):105–108.
- [9] **Forns, N., S. de Lorenzo, M. Manosas, K. Hayashi, J. M. Huguet, and F. Ritort.** 2011. *Biophys J* **100** (7):1765–1774.
- [10] **Hormeno, S., and J. R. Arias-Gonzalez.** 2006. *Biol Cell* **98** (12):679–695.
- [11] **Huguet, J. M., C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort.** 2010. *Proc Natl Acad Sci U S A* **107** (35):15431–15436.
- [12] **Landry, M. P., P. M. McCall, Z. Qi, and Y. R. Chemla.** 2009. *Biophys J* **97** (8):2128–2136.
- [13] **Leger, J. F., J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko.** 1998. *Proc Natl Acad Sci U S A* **95** (21):12295–12299.
- [14] **Liphardt, J., B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante.** 2001. *Science* **292** (5517):733–737.
- [15] **Maier, B., D. Bensimon, and V. Croquette.** 2000. *Proc Natl Acad Sci U S A* **97** (22):12002–12007.
- [16] **Manosas, M., D. Collin, and F. Ritort.** 2006. *Phys Rev Lett* **96** (21):218301.
- [17] **Manosas, M., I. Junier, and F. Ritort.** 2008. *Phys Rev E Stat Nonlin Soft Matter Phys* **78** (6 Pt 1):061925.
- [18] **Manosas, M., and F. Ritort.** 2005. *Biophys J* **88** (5):3224–3242.
- [19] **Melchionna, S., M. Fyta, E. Kaxiras, and S. Succi.** 2007. *Int. J. Mod. Phys. C* **18**:685.
- [20] **Meller, A., L. Nivon, E. Brandin, J. Golovchenko, and D. Branton.** 2000. *Proc Natl Acad Sci U S A* **97** (3):1079–1084.
- [21] **Mossa, A., M. Manosas, N. Forns, J. M. Huguet, and F. Ritort.** 2009. *J. Stat. Mech.* **P02060**.
- [22] **Ritort, F.** 2006. *J. Phys. Condens. Matter* **18**:R531–R583.
- [23] **Ritort, F., S. Mihadja, S. B. Smith, and C. Bustamante.** 2006. *Phys Rev Lett* **96** (11):118301.
- [24] **Rivetti, C., C. Walker, and C. Bustamante.** 1998. *J Mol Biol* **280** (1):41–59.
- [25] **SantaLucia, J. J.** 1998. *Proc. Natl. Acad. Sci. USA* **95**:1460–1465.
- [26] **Smith, D. E., S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante.** 2001. *Nature* **413** (6857):748–752.
- [27] **Smith, S. B., Y. Cui, and C. Bustamante.** 1996. *Science* **271** (5250):795–799.
- [28] **Smith, S. B., Y. Cui, and C. Bustamante.** 2003. *Methods in Enzymology* **361**:134–162.

- [29] **van Mameren, J., P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, and E. J. G. Peterman.** 2009. *Proc Natl Acad Sci U S A* **106** (43):18231–18236.
- [30] **Wagner, B., R. Tharmann, I. Haase, M. Fischer, and A. R. Bausch.** 2006. *Proc Natl Acad Sci U S A* **103** (38):13974–13978.
- [31] **Wang, M. D., H. Yin, R. Landick, J. Gelles, and S. M. Block.** 1997. *Biophys J* **72** (3):1335–1346.
- [32] **Wen, J.-D., L. Lancaster, C. Hodges, A.-C. Zeri, S. H. Yoshimura, H. F. Noller, C. Bustamante, and I. Tinoco.** 2008. *Nature* **452** (7187):598–603.
- [33] **Woodside, M. T., W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block.** 2006. *Proc Natl Acad Sci U S A* **103** (16):6190–6195.
- [34] **Wuite, G. J., S. B. Smith, M. Young, D. Keller, and C. Bustamante.** 2000. *Nature* **404** (6773):103–106.
- [35] **Yasuda, R., H. Noji, M. Yoshida, K. Kinosita, and H. Itoh.** 2001. *Nature* **410** (6831):898–904.
- [36] **Yin, H., M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles.** 1995. *Science* **270** (5242):1653–1657.
- [37] **Zuker, M.** 2003. *Nucleic Acids Res.* **31** (13):3406–3415. See also <http://mfold.rna.albany.edu/>.